This script is meant to use Linear Regression model

The dataset consists of 30 rows and 2 columns; where each row comprises of an employee and the 2 columns comprise of Years of Experience and Salary of each employee. The goal of this script is too use a Simple Linear Regression model to check and see if the independent variable (Years of Experience) for each employee can be used to predict the dependent variable (Salary)

1. Set working directory, import the libraries and the dataset

```
dataset = read.csv("Salary_data.csv")
```

2. Split the dataset into training and test set
load caTools library

```
library(caTools)
```

3. set seed

```
set.seed(123)
```

4. split the dependent variable where 2/3 goes to training set and rest to test

```
split = sample.split(dataset$Salary, SplitRatio = 2/3)
```

5. use the split to make training and test sets

```
training_set = subset(dataset, split == TRUE)
test_set = subset(dataset, split == FALSE)
```

6. Fitting Simple Linear Regression to the Training set

```
regressor = lm(formula = Salary ~ YearsExperience,data = training_set)
```

View the regressor

```
summary(regressor)
```

```
##
## Call:
## lm(formula = Salary ~ YearsExperience, data = training_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7325.1 -3814.4   427.7  3559.7  8884.6
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        25592       2646   9.672 1.49e-08 ***
## YearsExperience     9365        421  22.245 1.52e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 5391 on 18 degrees of freedom
## Multiple R-squared:  0.9649, Adjusted R-squared:  0.963
## F-statistic: 494.8 on 1 and 18 DF,  p-value: 1.524e-14
```

Predicting the Test set results

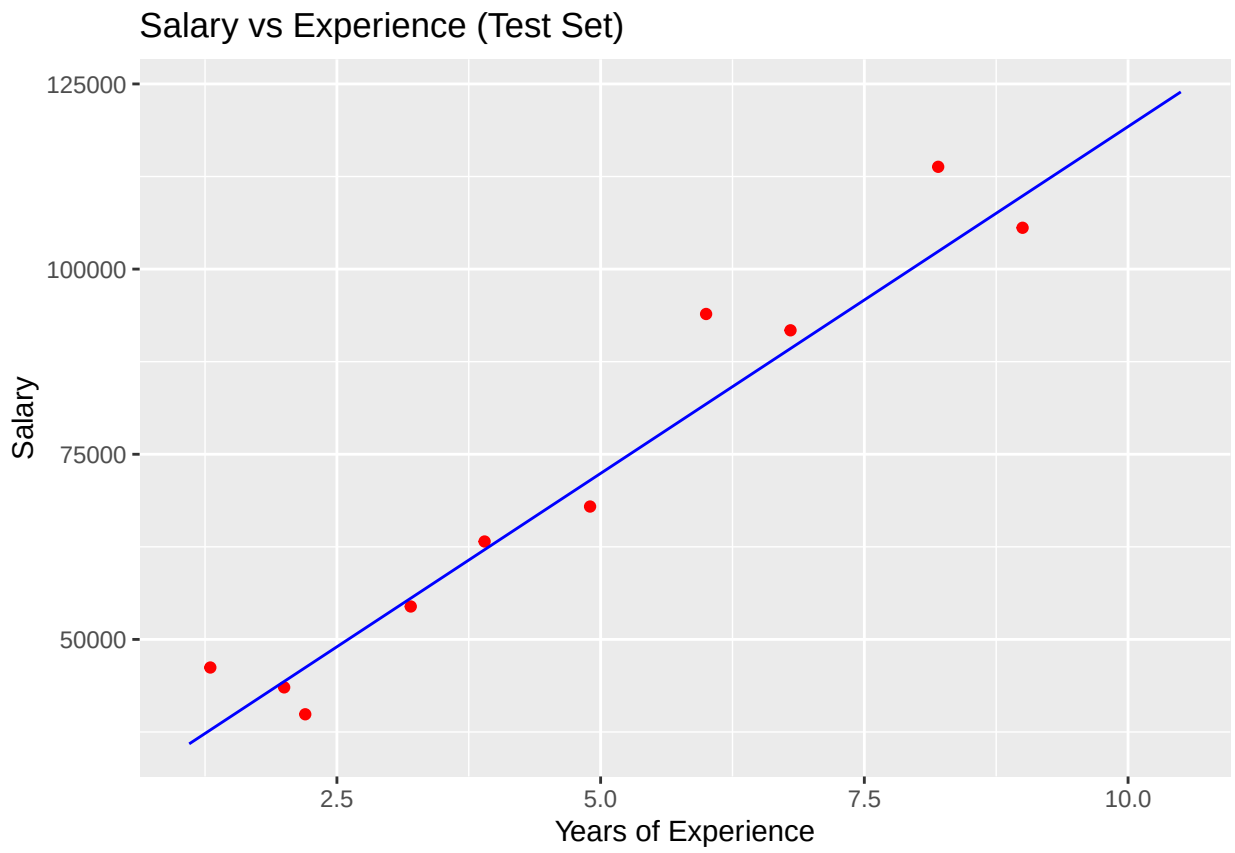
```
y_pred = predict(regressor, newdata = test_set)
```

7. Visualize the training set results

```
library(ggplot2)
ggplot() +
        geom_point(aes(x = training_set$YearsExperience,
                       y = training_set$Salary),
                   colour = 'red') +
        geom_line(aes(x = training_set$YearsExperience,
                      y = predict(regressor, newdata = training_set)),
                  colour = 'blue') +
        ggtitle('Salary vs Experience (Training Set)') +
        xlab('Years of Experience') +
        ylab('Salary')
```

8. Visualize the test set results

```
library(ggplot2)
ggplot() +
        geom_point(aes(x = test_set$YearsExperience,
                       y = test_set$Salary),
                   colour = 'red') +
        geom_line(aes(x = training_set$YearsExperience,
                      y = predict(regressor, newdata = training_set)),
                  colour = 'blue') +
        ggtitle('Salary vs Experience (Test Set)') +
        xlab('Years of Experience') +
        ylab('Salary')
```



‘‘